



Analysis of the use of open archives in the fields of mathematics and computer science

Anna Wojciechowska

► To cite this version:

Anna Wojciechowska. Analysis of the use of open archives in the fields of mathematics and computer science. OCLC Systems & Services, 2007, 23 (1), pp.54-69. sic_00131187

HAL Id: sic_00131187

https://archivesic.ccsd.cnrs.fr/sic_00131187v1

Submitted on 15 Feb 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analysis of the use of open archives in the fields of mathematics and computer science

Anna Wojciechowska*

* GERSIC : LVIC - University Aix-Marseille 3, France
annaw@cmi.univ-mrs.fr

Open access to Scientific and Technical Information (STI) is becoming more and more important. The Open Archives Initiative (OAI) has developed to the point that the principle is now accepted as an essential component in the communication of scientific results and in the publication of knowledge.

Such free access to STI can take two forms:

- self-archiving of articles, generally on the Web, and in particular in digital archives and repositories with public access (open archives),
- publication in open access (OA) journals.

In 2001, as part of a joint project with the CNRS¹ and the INRIA,² in partnership with the MathDoc³ unit, the CCSD⁴ developed “Hal”,⁵ a tool for scientific communication between researchers which, within the framework of the Open Archive Initiative (OAI), was aimed at promoting institutionalised self-archiving of research results in open access archives. For mathematics (and other disciplines), whenever the corresponding sub-discipline exists in “ArXiv”,⁶ all recent documents are automatically transferred from Hal into ArXiv.

The CNRS, Inserm,⁷ INRA⁸ and INRIA signed a joint declaration promoting open access to Scientific and Technical Information (Berlin, October, 2003). On March 22nd, 2005 they signed a common policy agreement to develop open archives with a common administrative and technical management.

The researcher is at the centre of the archiving system. This enquiry aims at studying the requirements and use of these open archives by researchers.

Presentation

A questionnaire was sent to part of the mathematical and computer science community in France by several libraries of the National Group of Mathematics Libraries. The fields (mathematics and computer science) were chosen because:

- there is currently much discussion about open archives in human and social sciences (drawing comparisons with what is done in science, technology and mathematics),
- these fields are of direct concern for the author of this report – a librarian in a mathematics and computer science library who wanted to check whether the library's users were aware of the existence of open archives, and if so, how they used them.

The first version of the questionnaire was tested in May 2005 with some researchers of the

¹ National Centre for Scientific Research, <<http://www.cnrs.fr/>>

² French Institute for Research in Computer Science, <<http://www.inria.fr/>>

³ National Centre for Documentary Coordination in Mathematics, <<http://www.mathdoc.ujf-grenoble.fr/>>

⁴ Center for Direct Scientific Communication (CCSD/CNRS), <<http://www.ccsd.cnrs.fr/>>

⁵ The Hal (“Hyper Article en Ligne”) archive provides authors with an interface enabling them to file manuscripts of scientific articles in every discipline in the database of the CNRS Center for Direct Scientific Communication (CCSD/CNRS), <<http://www.Hal.ccsd.cnrs.fr/>>

⁶ <<http://arxiv.org>> or <<http://ccsd.cnrs.fr/arXiv.html>>

⁷ French Institute for Health and Medical Research, <<http://www.inserm.fr/>>

⁸ French Institute for Agricultural Research <<http://www.inra.fr/>>

Laboratory of Analysis, Topology and Probability in Marseille (LATP). After several discussions some questions were reformulated or removed, others were added. Early in July 2005, the revised version was sent to the members of other laboratories in Marseille.

After some further corrections, the final version of the questionnaire was sent by e-mail in November 2005 to mathematicians and computer scientists via several libraries of the RNB⁹. The librarians of the RNB distributed this questionnaire to their users. The number of people reached by the questionnaire is estimated at 2200 ; of these, 128 persons answered.

Participation

Twelve research centers participated in this survey : Besançon, Bordeaux, Clermont Ferrand, Grenoble (Imag and the Fourier Institute), Marseille, Nancy, Paris (Jussieu and Orsay), Rouen, Strasbourg. The 128 persons who participated in the survey were essentially research lecturers and CNRS researchers, as shown in table 1 :

Table 1

<i>Participants</i>	<i>Percentage</i>
CNRS researcher	20.30%
Research lecturer	56.30%
Other	23.40%

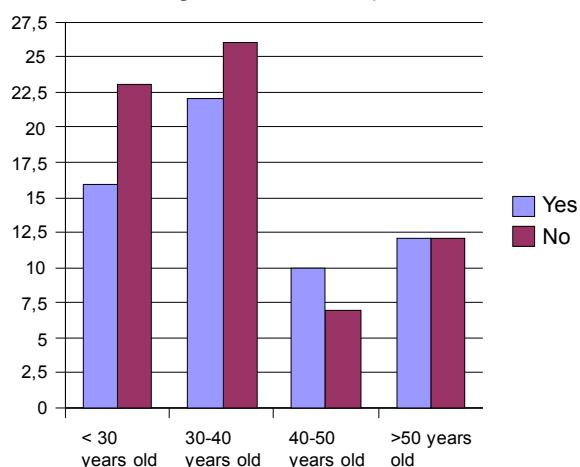
The productivity of a researcher is not the same at the beginning, middle or end of his career. Neither is the use of new technologies to access information. For this reason we included statistics concerning the age of the participants to try to reveal differences in use based on age. Almost 70% of participants were less than 40 years old. Participation in the investigation according to the age of the respondent is presented in table 2.

Table 2

<i>Age</i>	<i>Percentage</i>
< 30 years old	30.50%
30-40 years old	37.50%
40-50 years old	13.30%
>50 years old	18.70%

Almost half of the participants in the survey said that they knew the term “open archives” or had heard of these archives.

Knowledge of the term “open archive”



⁹ National Group of Libraries in Mathematics, <<http://www.rnbm.org/>>

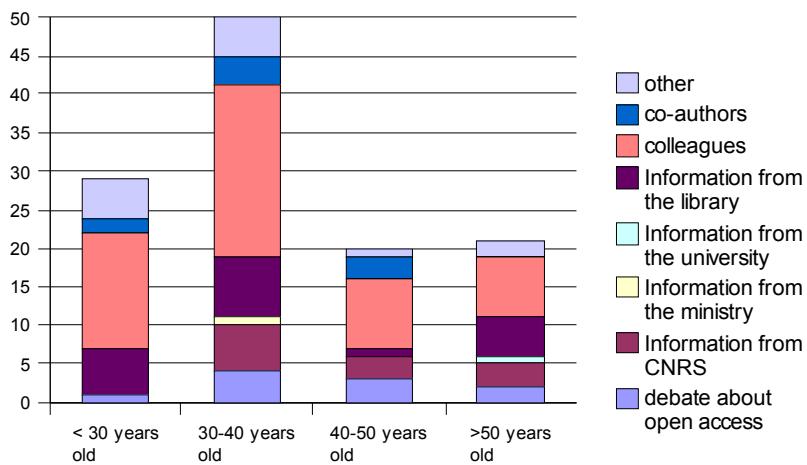
It seemed interesting to know how researchers learned of the existence of open archives, what motivated them to self-archive the articles, and what they knew about copyright and open access (OA) journals.

The probable sources of information on open archives are presented in table 3. Altogether 110 persons answered the question “How did you learn of the possibility of archiving your publications in institutional open archives?”

Table 3

<i>Source of information about open archives</i>	<i>Percentage</i>
colleagues	42.20%
information from the library	15.60%
other	10.00%
information from CNRS	9.40%
debate about open access	7.80%
co-authors	7.00%
information from the Ministry	0.80%
information from the university	0.80%

Source of information about open archives



“Colleagues” are a source of information about the existence of open institutional archives for 42% of researchers. “Other sources” includes this questionnaire and the newsletters of the INRIA.

Information search

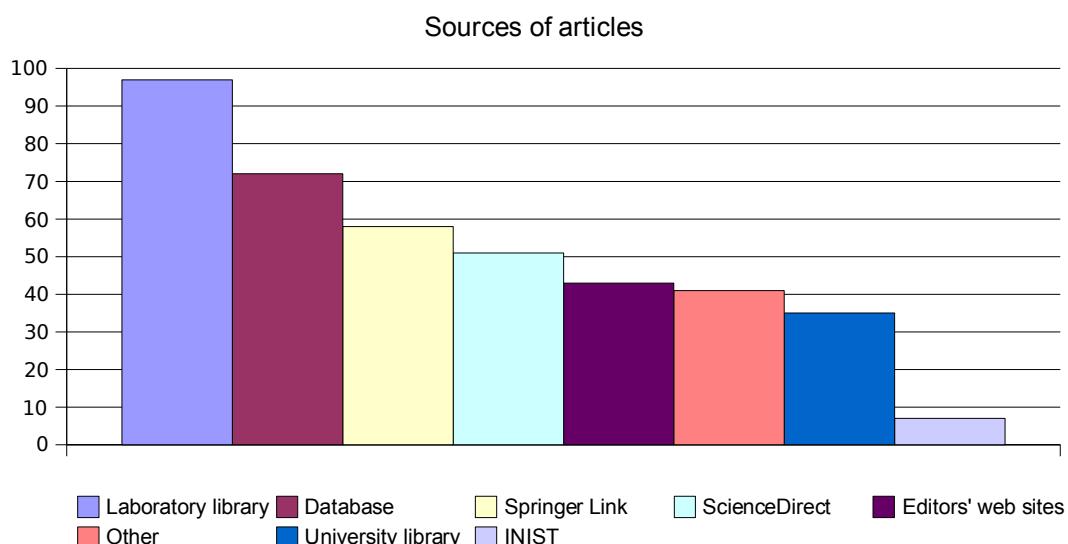
The researchers are both readers and authors of articles. For this reason, the first part of the questionnaire concerned researchers as readers and access to work-related information. This concerns especially the search for bibliographical references and for full text articles, both past and recent.

It seemed interesting to know where researchers find such information, how they search for it, whether they need assistance from librarians, how they use electronic information, how often they consult electronic articles, and how much recent the consulted articles are. To the question: “Where do you obtain the articles you need ?” the researchers could give several answers (table 4).

Table 4

<i>Source of articles</i>	<i>Percentage</i>
Department library	75.80%
Databases ¹⁰	56.30%
Springer Link ¹¹	45.30%
ScienceDirect ¹²	39.80%
Editors' websites	33.60%
Other	32.00%
University library	27.30%
INIST ¹³	5.50%

Even if 76% of respondents find the articles (or their references) in their department's library, bibliographic databases are quite heavily used. 56% of respondents have already used bibliographical databases as a source of information, and journals with full text on line (with subscription paid by the library or university) are consulted more and more : 45% of researchers find the articles in Springer Link and 40% in ScienceDirect. 32% give other sources of articles, like personal Web pages, ArXiv, Hal, other on-line subscriptions like Jstor,¹⁴ CiteSeer, or direct contact with the authors.



When it comes to searching the Web for research articles with open access to the full text, a majority of researchers use Google (66%) and ArXiv (66%). Some of them use Emis,¹⁵ personal pages or Numdam.¹⁶

Google indeed makes it possible to find personal Web pages, the pages of the laboratories or of the libraries locally storing scientific publications, and it also enables direct access to full text articles. The use of ArXiv means access to e-prints in particular, though some published articles can also be found there.

¹⁰ MathSciNet (<<http://e-math.ams.org/mathscinet/>>) and Zentralblatt (<<http://www.emis.de/ZMATH/>>), subscription access paid for by libraries.

¹¹ SpringerLink : Springer, Birkhauser et Kluwer' subscription access to full text of journals, <<http://www.springerlink.com>>

¹² Elsevier's journals in full text, payed by universities <<http://www.sciencedirect.com>>

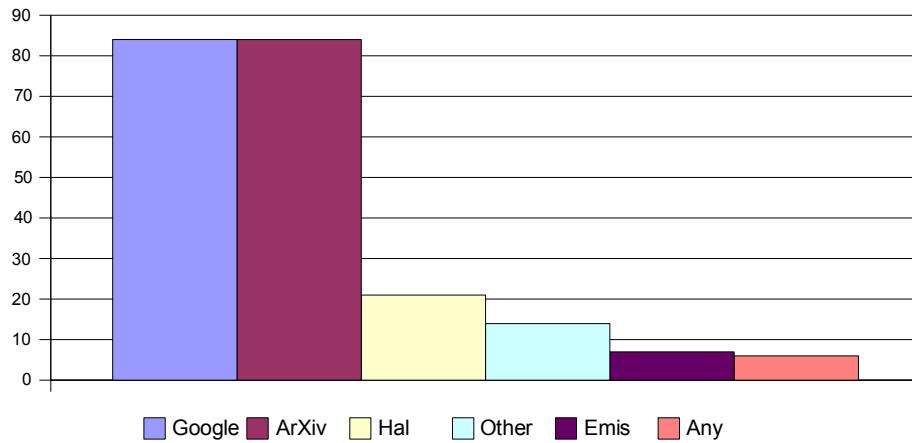
¹³ INIST : National Institute for Scientific and Technical Information, <<http://www.inist.fr/>>

¹⁴ The Scholarly Journal Archive : <<http://www.jstor.org/>>

¹⁵ This site lists electronic journals and hosts home pages or mirrors of many of them, <<http://www.emis.de/journals/>>

¹⁶ Folder database of digitized mathematics, <<http://www.numdam.org/>>

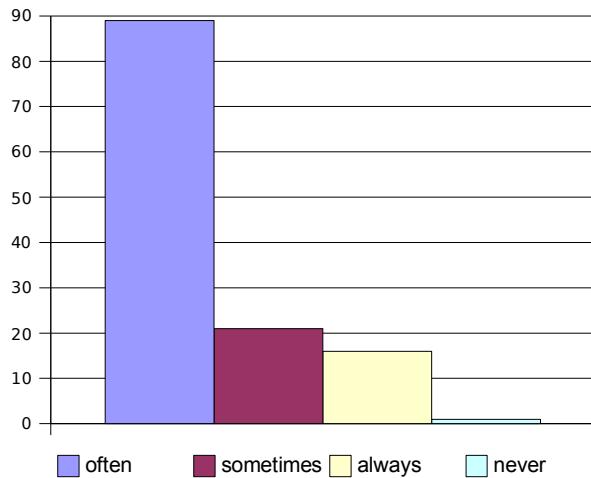
Access to the full text



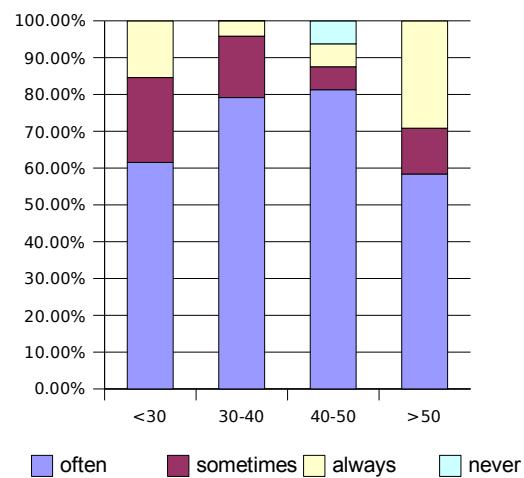
Answers to the question concerning access points to e-preprints confirmed familiarity with Hal and ArXiv (for 58%), but 77% of researchers access electronic preprints via personal pages. Access via laboratory or library sites was indicated by 18.8%, which means nearly 1 researcher out of 5 knows that a local archive exists and is able to access it directly.

More than 80% of the respondents find the articles they need without any difficulty, and this independently of the age of the researcher.

Access to articles



Access to articles by age



Researchers usually ask librarians to find the references or publications they need. But the possibility of finding information on line has made this easier. So it seemed interesting to know whether these articles are still accessed with the assistance of the librarians (table 6).

Table 6

<i>Ask for help</i>	<i>Percentage</i>
always	0.00%
often	5.50%
sometimes	58.60%
never	35.20%

Almost 60% ask for the librarian's help only sometimes, and 35% do not need any help. Full text articles are recent (usually after 1995). But more and more older articles digitized within the framework of various local, national or international projects are also on-line in open access.

If researchers do not need help in retrieving information, it is because almost 40% of them use articles which were published during the last 10 years, i.e. articles which are for the most part available either on line in full text, or in paper version in libraries.

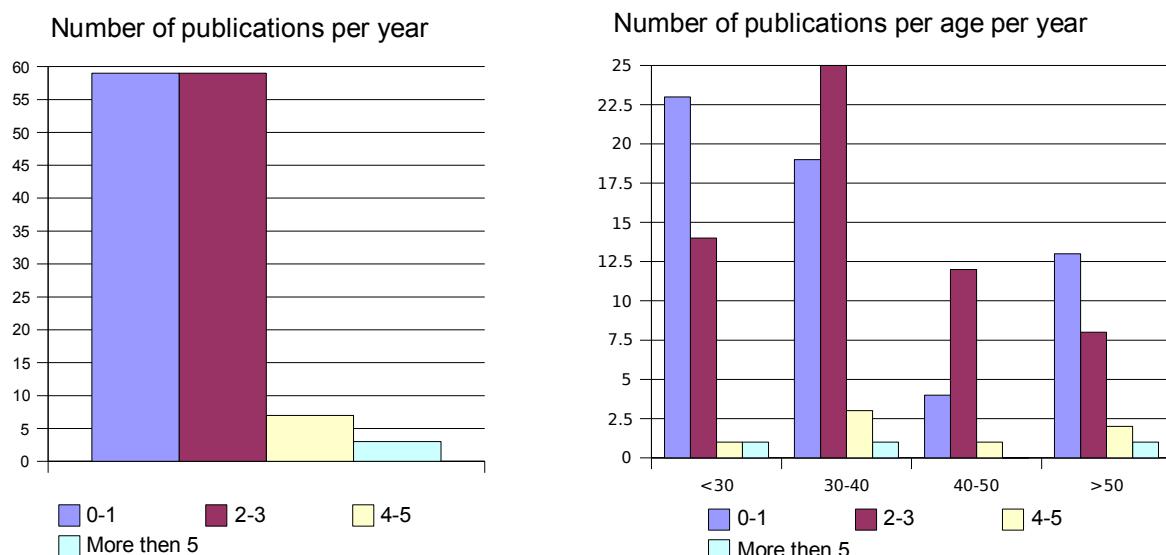
E-publications are consulted by almost 60% of those questioned once per week or more. Almost all (93%) consult preprints on line, but, as we shall see later, not so many deposit their preprints on line in open access.

Publications

This part analyses the amount of researchers' publications available, how they are filed in open archives, the types of filed articles, etc.

A publication can take the form of a preprint, i.e. of a written text without peer review or editorial review, or of postprint – the final published version of article.

46% of questioned researchers publish a maximum of 1 article per year, but these are mostly recent PhDs or young lecturers. 46% of those questioned said they publish 2-3 articles per year.



The publication of an article generally corresponds to an advance in a research project. It is on the basis of these articles that the researcher is evaluated and financed. For those researchers who answered the questionnaire, the main reason for publishing their work was to communicate their research findings to the scientific community (table 7).

Table 7

<i>Aim of publication</i>	<i>Percentage</i>
Communication of results to the community	86.00%
Career advancement	37.50%
Personal prestige in the domain	17.20%
Increased possibilities for future funding	9.40%
Other	5.50%

Each of them could give several answers. Among other answers there is scientific curiosity, the wish for the results to be recognised and accepted, the wish to clearly formulate the results obtained.

Experience in self-archiving

Self-archiving consists in posting an electronic document (generally an article, for a researcher) on a website, a document which can then be consulted freely by all. This is done in order to maximize visibility of, and accessibility to, the results.

There are three ways a researcher can provide open access to articles:

- he can deposit a copy of an article on a personal or institutional website,
- he can place it in an institutional open access archive (such as Hal),
- he can put it in a disciplinary open access archive (such as ArXiv).

Scientific publications are mainly posted online by the authors and/or the co-authors (table 8). But some of the authors who publish between 2 and 4 articles per year do not archive them (the articles are sent directly to the publishers). The documents posted by the administrative staff of the laboratories usually do not contain full text. They are merely postings of bibliographical notes.

Table 8

<i>Author of deposit</i>	<i>Percentage</i>
You	74.20%
Your co-author	10.90%
Staff of your laboratory	11.70%

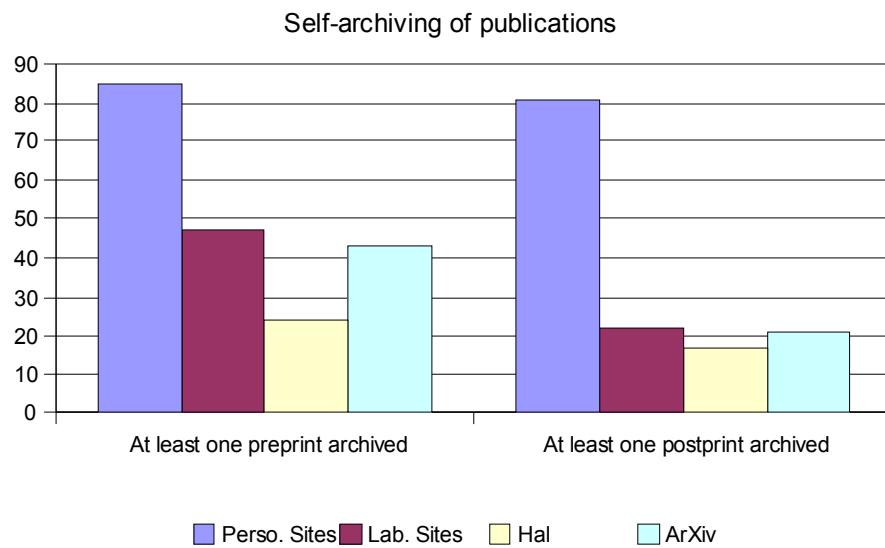
To the question “What kind of documents do you put into open archives ?” (table 9), 65% of those questioned posted postprints, and 77% preprints. Among other types of publications we find errata and files of conference talks (e.g. Powerpoint files).

Table 9

<i>Types of publication deposited</i>	<i>Percentage</i>
refereed article	64.80%
congress paper	22.70%
preprint	77.30%
technical report	13.30%
book chapter	6.00%
dissertation, thesis etc.	23.40%
course notes	18.00%
exercises	18.00%
other	2.30%

73% do not fear “plundering” or improper use of the e-prints, which corresponds (at least) to the number of depositors of preprints in open archives. Among those who archive their publications in institutional open archives, 55% do so as a matter of principle, and 25% because they're there.

To the question: “How many articles did you deposit in open archives over the last 3 years”, some researchers declared having posted articles on personal websites. Yet, among the researchers who did not deposit any article, most did not have personal Web pages. This question was divided into several sub-questions, making it possible to separately evaluate postings of preprints and articles with peer review into personal and institutional sites.



Some people did not answer these questions : either the posting was made by the co-author/the administrative staff, or quite simply the researcher did not give an answer (see the table 8).

Self-archiving respondents were asked what their original motivation was for self-archiving their work. Researchers could give several answers, which means that probably those which self-archived articles on their websites also did so (at least partly) in institutional archives. Thus, the total percentage do not indicate the true level of self-archiving, especially as the articles could be deposited by other people (co-authors, etc.).

Here below are the details about self-archiving :

Preprints on personal websites

66.4% of respondents posted at least 1 preprint on a personal site.

Among the 29.7% of those which did not post anything, the majority consists of PhD students, and researchers over 50.

Preprints on laboratory or library websites

36.7% of respondents posted at least 1 preprint on the site of the laboratory or the library.

Preprints on Hal

18.6% of respondents filed at least 1 preprint on Hal.

Among them, 87.5% filed at least 2 preprints on Hal.

Preprints on ArXiv

33.6% people filed at least 1 preprint on ArXiv.

Among them, 79% filed at least 2 preprints.

Postprints on personal websites

63.3% of respondents posted at least 1 refereed article on a personal website.

Postprints on laboratory or library websites

17.2% of respondents posted at least 1 refereed article on the site of the laboratory or the library.

Postprints on Hal

13.3% of respondents filed at least 1 refereed article on Hal.

Among them, 82.4% deposited at least 2 refereed articles.

Postprints on ArXiv

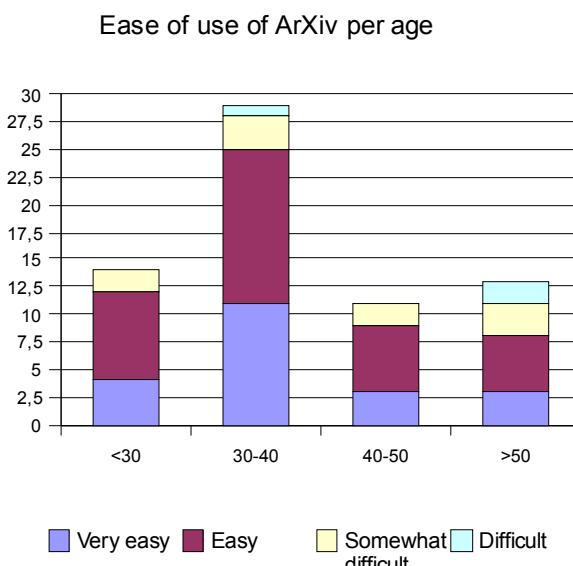
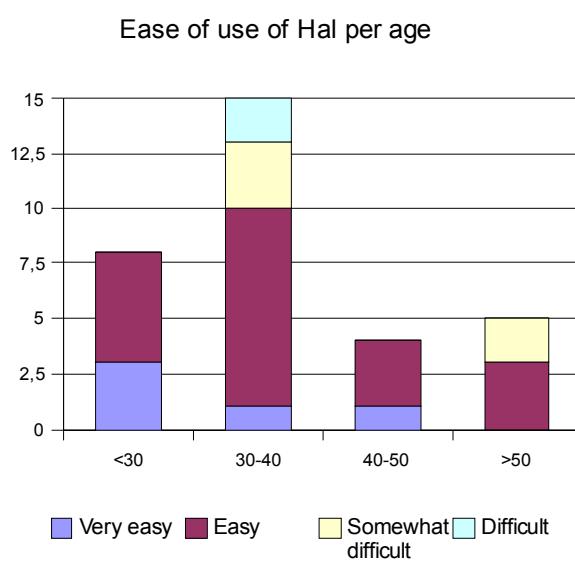
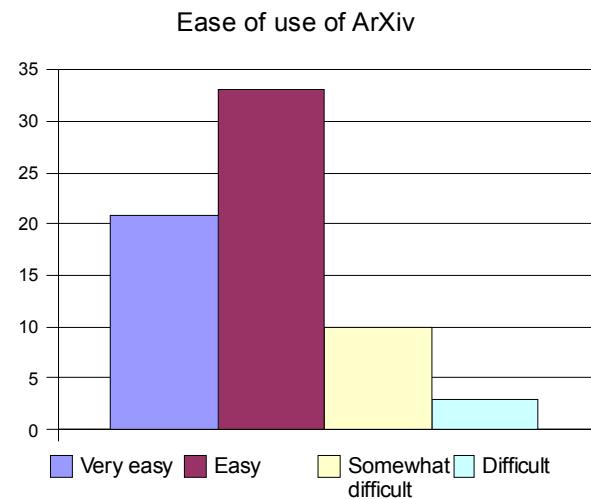
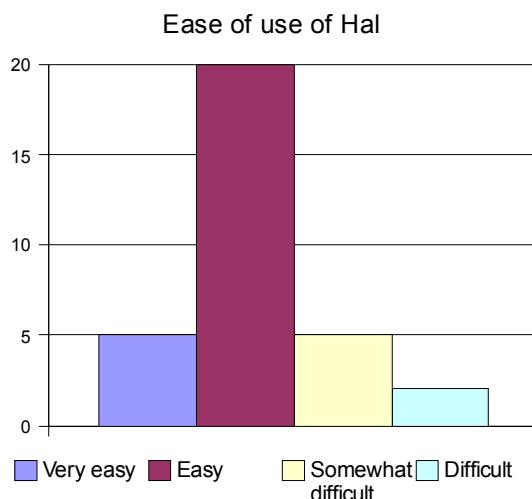
16.4% of respondents filed at least 1 refereed article on ArXiv.

Among them, 71.4% filed at least 2 refereed articles.

Answers were given by only 128 people, but a look at the Hal site shows that there are fewer publications in mathematics and computer science than in physics for example.¹⁷

Then, it seemed interesting to determine the opinion of researchers on the ergonomics of Hal and ArXiv and about the procedure of self-archiving. 124 researchers gave an answer on the use of Hal and 127 on the use of ArXiv.

The findings here show that among those who gave an opinion on the use of Hal, 62.5% find it easy. Among those which gave an opinion on the use of ArXiv, almost 50% find it easy and 31.3% very easy. Among the “no opinion” answers, there is a certain number of people who don't deposit their articles.



¹⁷ As of 31/12/05 the number of publications filed on HAL was as follows : physics : 8479, mathematics : 2181, computing : 1736

The next questions concerned the necessary time to self-archive an article. To complete a deposit on Hal, the first deposit takes less than 30 minutes for 34.5% of respondents and subsequent deposits less than 15 minutes for 81.5% of them. To complete a deposit on ArXiv, the first deposit takes less than 30 minutes for 41.3% of respondents and subsequent deposits less than 15 minutes for 73.3%.

The directors of the different scientific departments of the CNRS asked the laboratories to deposit a copy of all their published articles in an institutional repository – Hal –, but the CNRS did not make this compulsory. The question “What would be your reaction if your employer (CNRS or Ministry of Education) required you to deposit your publications in an institutional repository?” was voluntarily provocative and the vast majority (74%) of respondents said they would comply willingly with this requirement. Only 7.80% said they would comply reluctantly and 6.20% said they would not do so.

Copyright awareness

An author can file in open archives any type of document of which he is the intellectual owner. This concerns documents already published or in press, documents in the course of scientific validation (preprints) or work papers.

Authors can make digital postprints available when :

- they have retained copyright and granted only non-exclusive rights to the publisher,
- they have transferred all rights to the publisher, but the publisher's policy permits authors to distribute postprints under specified terms and conditions (most publishers now have such self-archiving policies),
- they have modified the preprint using errata/corrigenda.

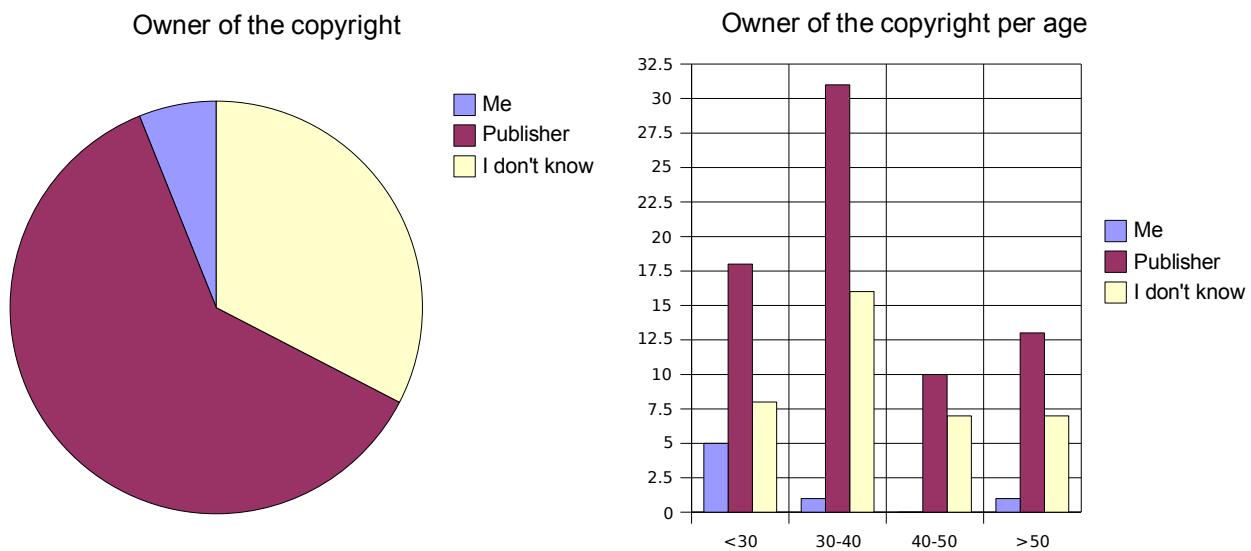
Only explicit prohibition of copyright transfer in a contract (which then gives exclusively to the publisher the rights to electronically exploit the document) makes it compulsory for the author to ask for the publisher's permission to file a document in open archives.

It seemed interesting to find out what researchers really knew concerning the legal aspects of scientific publications and to ascertain whether researchers read the copyright statements provided by publishers.

Authors were asked who retained the copyright to the last article they had self-archived. 56.2% said that it remained with the publisher, 30% did not know it. Only 5.5% said it was themselves.

Of the researchers who had self-archived (especially on personal Web pages) their last published article, a majority (54%) did not know whether the publisher's permission was necessary before doing so. 18% didn't ask their publisher for it, and only 7.8% did ask. 5.5% said they had copyright on their last published article. Besides, 77.3% did not know that they could negotiate self-archiving with their publisher.

These answers can mean that the self-archiving was not prohibited in the contracts of copyright transfer, or that the articles were published without the publisher's permission. It is possible to think here that a majority of researchers sign the copyright statements provided by publishers without reading them.



Knowledge of open access (OA) journals

Publication in open access journals is the second form of open access to scientific information. Questioned on whether they had submitted a manuscript to an open access journal in the last three years, only 17% gave a positive answer.

Researchers were invited to indicate the reasons for publishing their work in an open access journal. The authors who publish there (22 respondents) are generally motivated by the principle of open access and the good reputation of this journal in their field.

People who did not publish in open access journal were asked to indicate the reasons for not having done so, and the results are as follows (the percentage was calculated with respect to the 100 people who said they had not published in open access journals) :

- the researchers do not know OA journals in their field (72%),
- they say that the open access journal in their field are not prestigious (10%),
- they are against the author-payer principle (7 %),
- Other reasons : article rejected.

Finally, the intentions of researchers are as follows: 34.4% of researchers plan to publish in an OA journal in the future and 38.3% do not know yet if they will do so.

Yet the level of awareness of open access journals is relatively low in mathematics, and there are more than 80 titles (<http://www.doaj.org/>). Moreover, the “author-payer” economic model, which exists in other countries, is not popular in France, and it is probably not well known.¹⁸

Conclusions

- A majority of researchers find the articles (or their references) necessary for their work in libraries, but journals with full text (with subscriptions paid by the libraries) are often consulted.
- More than 80% of respondents find the articles they need without any difficulty. Almost 60% of them ask for the librarian's help only sometimes and 35% do not need any help.
- To obtain articles with full text in open access, the researchers use especially ArXiv (to get electronic preprints) and Google (to search for personal Web pages).
- The most consulted (at least once a week) full text articles available on-line were published

¹⁸ For instance, the new Springer-Kluwer group has launched Open Choice : if an author chooses to pay (\$ 3500), his article will be available without a subscription being necessary.

during the last ten years.

- Almost 50% of researchers publish 2-3 articles per year.
- A majority of them post a copy of their articles on personal websites, and 28% have done so for at least 5 years.
- Part of those researchers who post publications on personal sites also simultaneously file them in institutional open archives. But researchers always post many more articles on their personal webpages (63%) than in Hal (12%) or in ArXiv (16%).
- Those who have already filed publications in Hal or ArXiv find them easy to use and say they needed less than 30 minutes to file their first article and less than 15 minutes for subsequent filings.
- The authors claim that they are becoming more aware of open access, and that they learned of the existence of institutional open archives thanks to communication between them.
- Yet the sources of on-line articles with open access, such as OA journals, are still not very well-known.
- Official information coming from the CNRS or the Ministry was on the whole unnoticed.
- A majority of researchers would accept being required to file publications in institutional open archives.
- Those who are already filing publications in open archives do so as a matter of principle.
- A majority of them do not read the contracts they sign with their publishers, and they do not know of the possibility of negotiating these contracts.
- Most researchers sign the copyright statements provided by their publishers without reading them.

With regard to the self-archiving of articles in open institutional archives:

To increase the number of publications filed in institutional archives, it would be necessary to encourage the researchers and help them to adopt this new style of publication in order to improve the distribution of their scientific production.

The difficulties in the development of open archives are not technical, but social. The practices of self-archiving already form part of the work pattern of researchers in mathematics and computer science, but only as a kind of "Google" reflex, i.e. depositing publications on personal sites to ensure that they can be found by the most commonly used search engine. The utility of institutional open archives is not yet well understood.

The development of open archives is founded on the self-archiving of scientific publications by their authors. It would be necessary to envisage training researchers to the use of Hal. Making the legal aspects of scientific publication more widely known, and sensitizing researchers to the necessity of checking the contracts they sign with their publishers, also appear to be key issues.

With regard to the publication of OA journals:

According to a study undertaken by the Center for Information Behaviour and the Evaluation of Research (CIBER),¹⁹ more and more scientists publish in OA journals. Yet it's impossible to say so of French researchers in mathematics and computer science for the moment. It is necessary to increase the dissemination of information about open journals since, according to the results of this questionnaire, they are not sufficiently exploited. It could be the role of libraries to circulate relevant information from the publishers of OA journals.

¹⁹ <http://www.ucl.ac.uk/ciber/ciber_2005_survey_final.pdf>

Bibliography

Allen, James. Interdisciplinary differences in attitudes towards deposit in institutional repositories (2005). Available at <http://eprints.rclis.org/archive/00005180/>

Aubry C., Janik J. (dir.). 2005. Les archives ouvertes : enjeux et pratiques. Guide à l'usage des professionnels de l'information. Paris: ADBS, 332 p.

Centre for Information Behaviour and the Evaluation of Research (CIBER). Available at http://www.ucl.ac.uk/ciber/ciber_2005_survey_final.pdf

Chanier, T. 2005. Archives ouvertes et publication scientifique. Comment mettre en place l'accès libre aux résultats de la recherche ? L'Harmattan, Paris, 188p.

Fily, Marie-Françoise. 2005. Introduction au concept d'archive ouverte. Available at http://archivesic.ccsd.cnrs.fr/sic_00001523.html

Gallezot, Gabriel. Le Libre Accès (Open Access) : partager les résultats de la recherche, Colloque international : L'information numérique et les enjeux de la société de l'Information - Tunis, 14-16 Avril 2005 - ISD. 14 avril 2005. Available at http://archivesic.ccsd.cnrs.fr/sic_00001416.html

Guide juridique du CNRS, en cours de publication. Available at <http://publicnrs.inist.fr/>

JISC Disciplinary Differences Rapport (2005). Available at http://www.jisc.ac.uk/uploaded_documents/Disciplinary_Differences_and_Needs.doc

Libre accès à l'information scientifique et technique. Available at <http://www.inist.fr/openaccess>